Clearing your Desk! Software and Data Services for Collaborative Web Based GIS Analysis

David Tarboton, Ray Idaszak, Jeffery Horsburgh, Dan Ames, Jon Goodall, Larry Band, Venkatesh Merwade, Alva Couch, Rick Hooper, David Maidment, Pabitra Dash, Michael Stealey, Hong Yi, Tian Gan, Tony Castronova, Brian Miles, Cuyler Frisby, Zhiyu Li

USU, RENCI, BYU, UNC, UVA, CUAHSI, Tufts, Texas, Purdue, Caktus

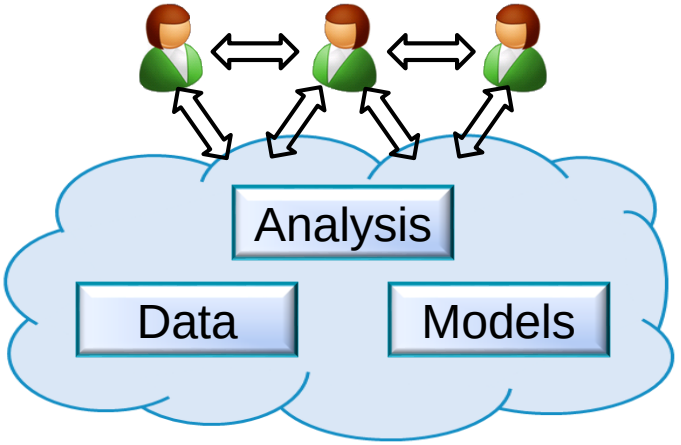
http://www.hydroshare.org

HYDROSHARE

OCI-1148453
OCI-1148090
2012-2017

Outline

- Data and computational challenges
- HydroShare
  - Goals
  - Resource data model
  - Architecture
- Terrain analysis and TauDEM in OpenTopography and CyberGIS
- Data services for hydrologic modeling
- Summary



The challenge of increasing Digital Elevation Model (DEM) resolution

1980's DMA 90 m

10^2 cells/km^2

1990's USGS DEM 30 m

10^3 cells/km^2

2000's NED 10 m

10^4 cells/km^2

2010's LIDAR ~1 m

10^6 cells/km^2

e.g. 50,000 km^2 Watershed

27 MB

240 MB

2 GB

200 GB

Data Heterogeneity

- From dispersed federal agencies
- From investigators collected for different purposes
- Different formats
  - Points
  - Lines
  - Polygons
  - Fields
  - Time Series

- The way that data is stored can enhance or inhibit the analysis that can be done
- We need ways to organize the data we work with
- Data models

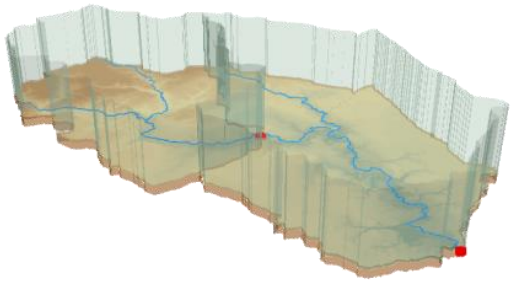
Water quality

Water quantity

Rainfall and Meteorology

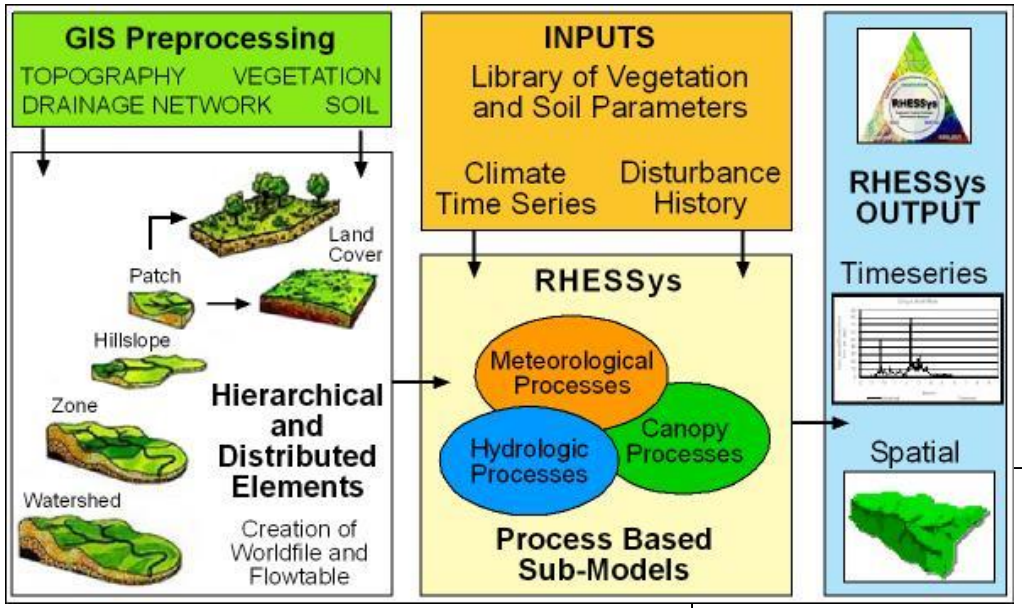
Soil water

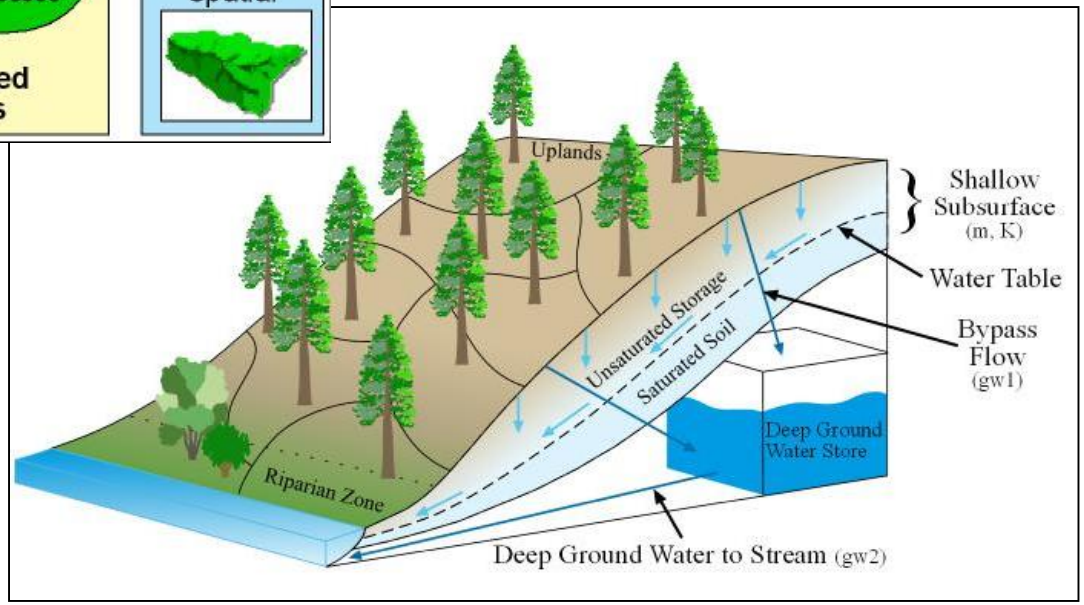
GIS



Groundwater

Data intensive models to understand and examine consequences, impacts and effects of land surface and climate changes

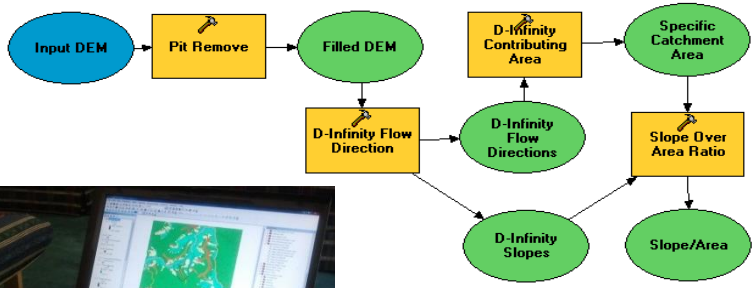




From Larry Band

Do you have the access or know how to take advantage of advanced computing capability?

Hydrologic Experimentation and Modeling



Data Intensive High Performance Computing

#!/bin/bash vi chmod

#PBS -l nodes=4:ppn=8

grep awk mpiexec

```
-bash-3.2$ ls tddata
logan                LoganOutlet.sbn  LoganOutlet.shp      LoganOutlet.shx
LoganOutlet.dbf  LoganOutlet.sbx  LoganOutlet.shp.xml
-bash-3.2$ ls tddata/logan
logan.tif
-bash-3.2$ ls
eric    logMFfel        run.bash        taudem.bash     taudem_submit.sh
logMF   run_all.bash    run_taudem.sh   taudem.o41959   tddata
-bash-3.2$ run_taudem.sh pitremove -z logan -fel loganfel
43058.ib-net
-bash-3.2$
```

A digital divide

Data and Software Services

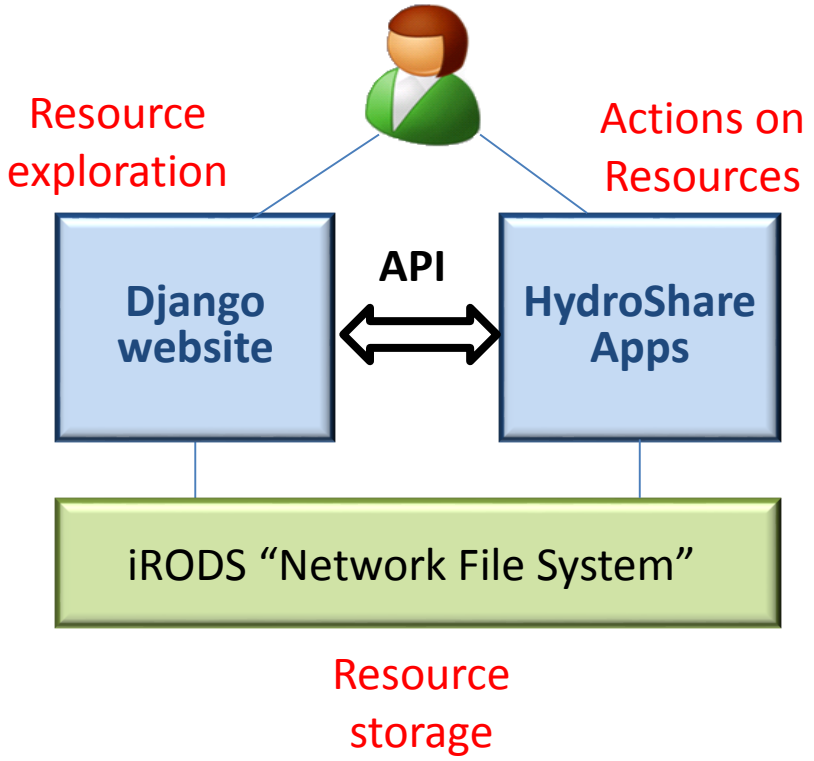
HydroShare Goals

- To provide a cyberinfrastructure platform for hydrologic research to solve problems of size and scope not otherwise solvable using desktop computing through
  - Software as a service
  - Data as a service
  - Models as a service
  - Visualization and analysis services
- To enable more rapid advances in hydrologic understanding through collaborative data sharing, analysis and modeling
- To address community cyberinfrastructure needs

CUAHSI
universities allied for water research

HydroShare is a collaborative environment (being developed) for data sharing, analysis and modeling

- Share your data and models with colleagues
- Manage who has access to the content that you share
- Share, access, visualize and manipulate a broad set of hydrologic data types
- Sharing and execution of models
- Web services API to facilitate automated and client access to almost all functionality
- Access to and use of high performance computing
- Publication of data and models with a DOI



Our goal is to make sharing of hydrologic data and models as easy as sharing videos on YouTube or shopping on Amazon.

Functionality

- Sharing and publication of data
- Social discovery and added value
- Model sharing

Collaboration

Server/Cloud Computation

- Model input data preparation
- Model execution
- Visualization and analysis (best of practice tools)

- Platform independence
- Big data
- Reproducibility
- Software installation and configuration

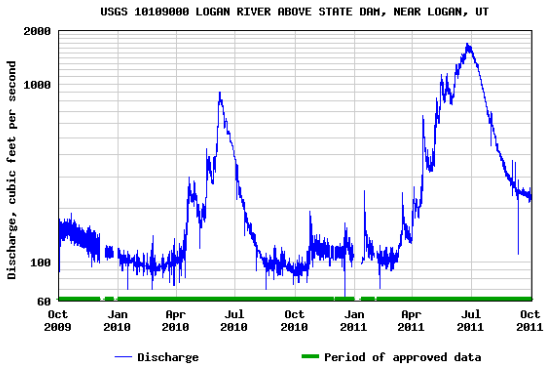
HydroShare is a system for sharing Resources and Collaborating

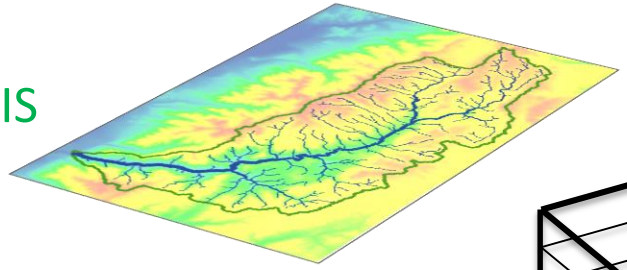
- Files and sets of files structured to represent a hydrologic process, model, or element in the hydrologic environment
- Standard data models enhance interoperability and support functionality “hydro value added”
- Tools that act on resources to visualize, modify and create new resources
  - Encode standard/best practices
- Access control and sharing model

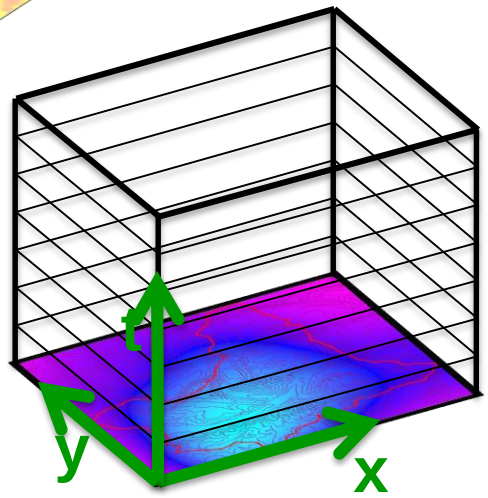
Types of data to support as resources

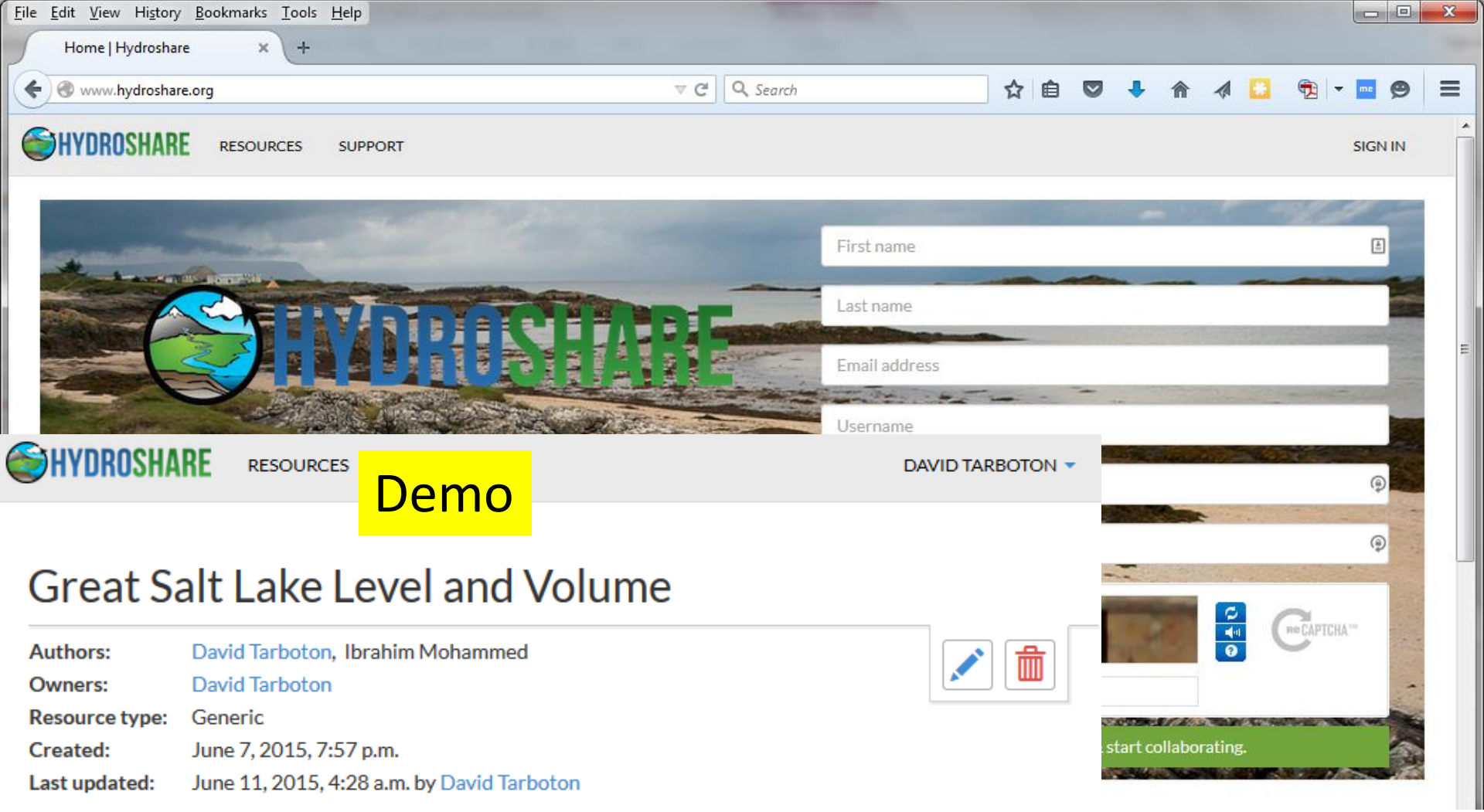
Resource Types

- Generic ✓
- Geographic Raster ✓
- Time Series ✓
- Multidimensional Space Time dataset ✓
- Model program ✓
- Model instance ✓
- Geographic Feature set ✓
- Referenced Time Series (CUAHSI HIS web service link) ⚠
- Application ⚠
- River Geometry
- Sample based observations (ODM2 and CZO)
- Model component
- Composite resources









HYDROSHARE RESOURCES Demo DAVID TARBOTON

Great Salt Lake Level and Volume

Authors: David Tarboton, Ibrahim Mohammed
Owners: David Tarboton
Resource type: Generic
Created: June 7, 2015, 7:57 p.m.
Last updated: June 11, 2015, 4:28 a.m. by David Tarboton

Abstract

These comma separated variable files give the level and volume of the Great Sa
1847 to 2014-05-03. Level in feet is as recorded by the USGS. Level in m was c
were computed from the bathymetry.

How to cite

Tarboton, D., I.Mohammed (2015). Great Salt Lake Level and Volume, HydroSh
/resource/7875d551e40a43b8848f74c63f5481ae

Content

data/contents/GSLN2013.txt
608.4 KB

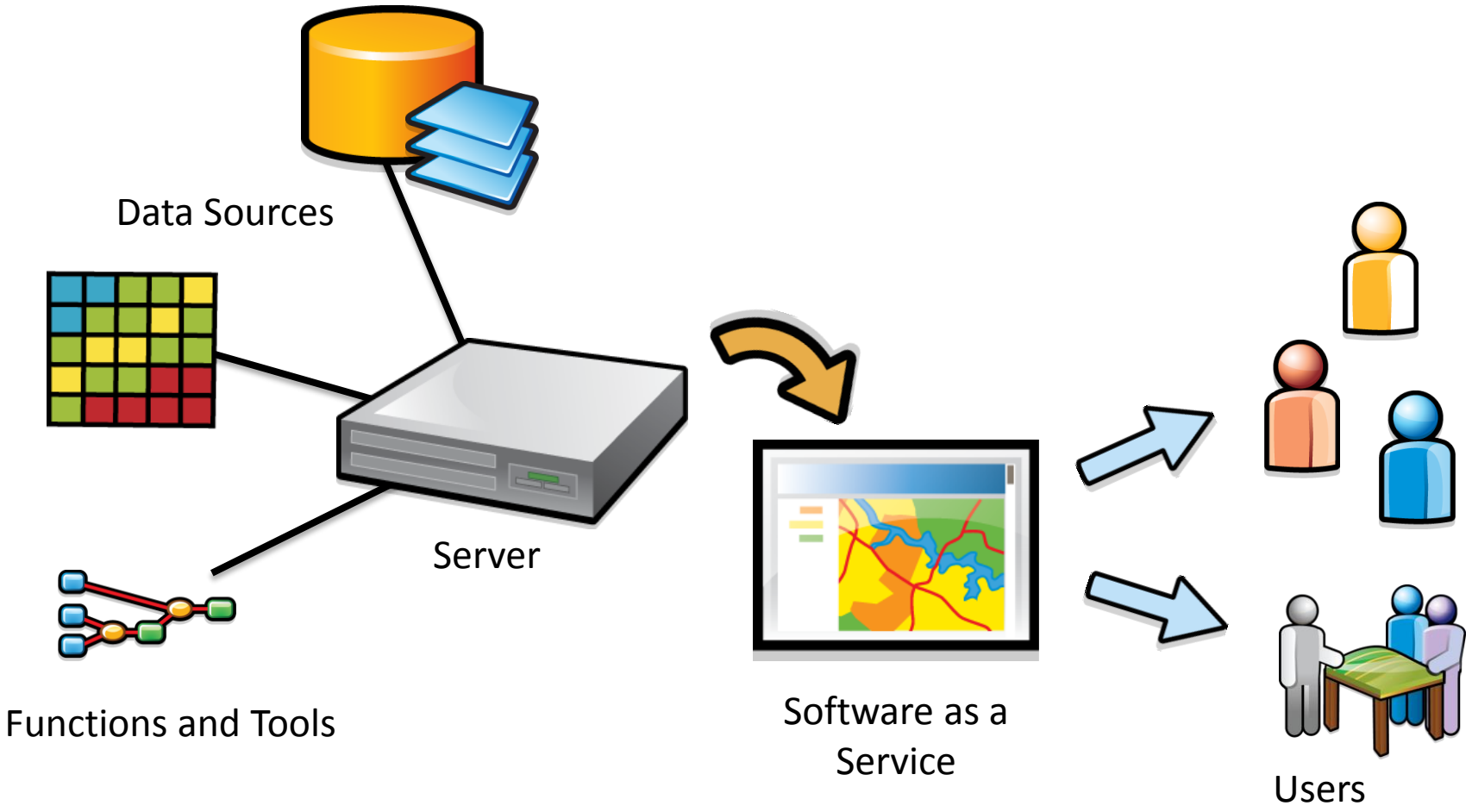
data/contents/GSLS2013.txt
681.9 KB

Download All Content as Zipped BagIt Archive | + Add file... | + Sign in iRODS...

Learn more about the Bagit archive format

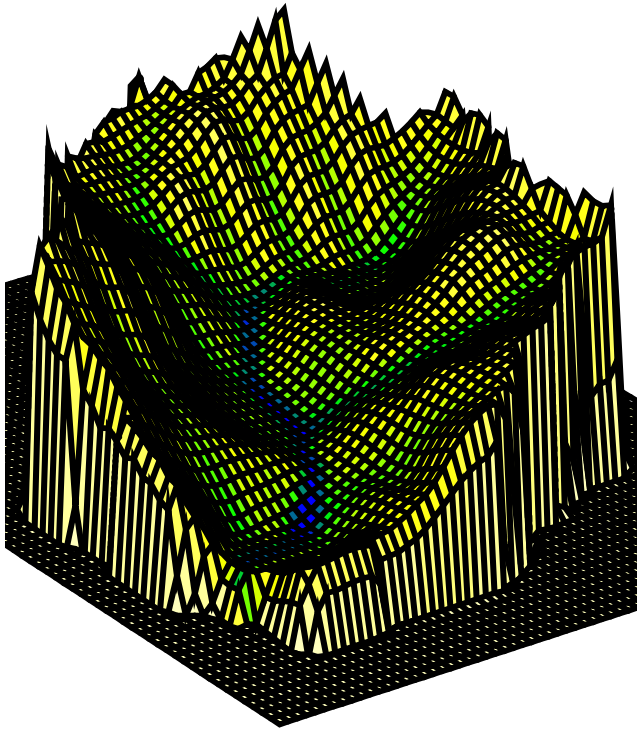
Clearing your desk. The trend towards network (cloud) computing.

Can we deliver GIS and Hydrologic Analysis functionality as services over the web?

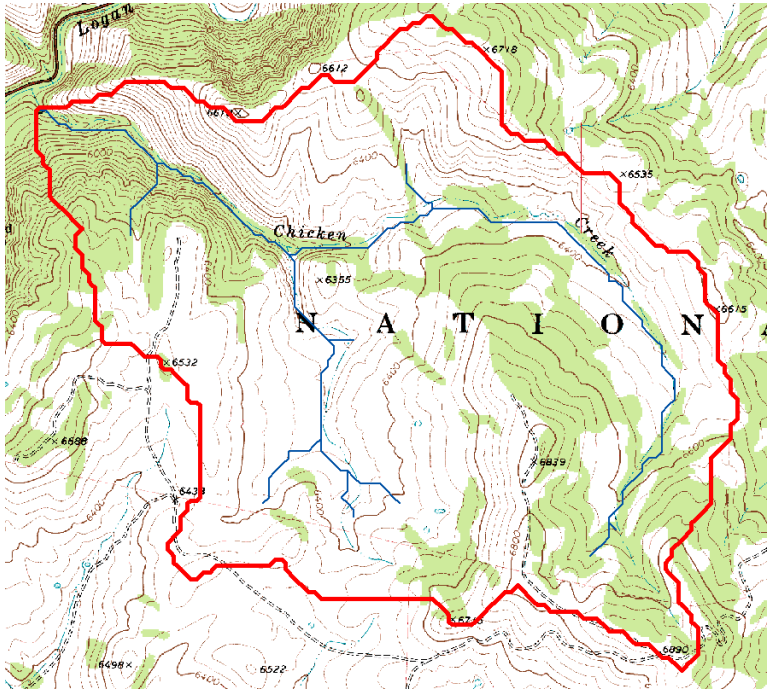


Based on slide from Norm Jones

Terrain Analysis

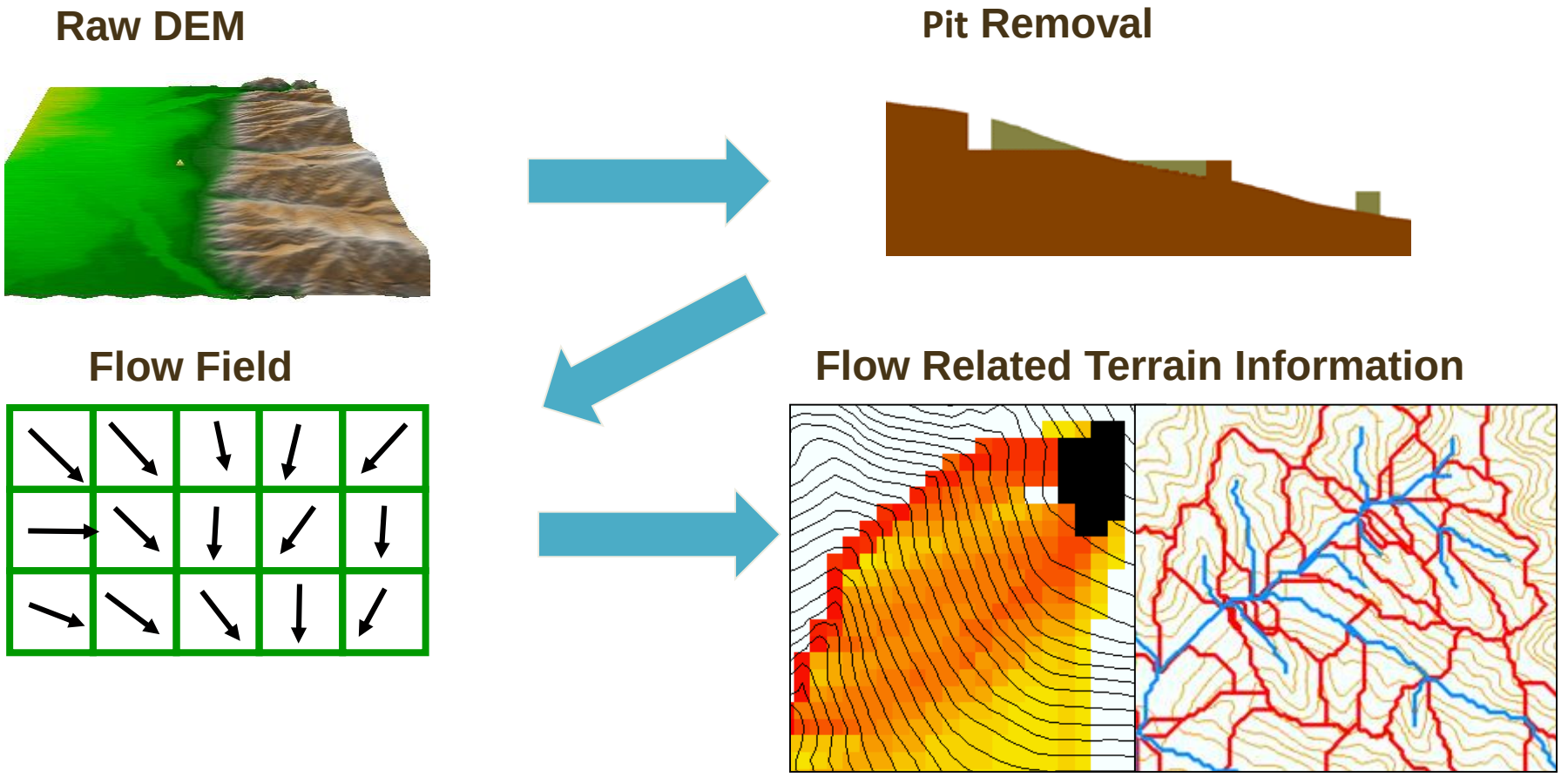


- Topography is fundamental to hydrology
- Watersheds are the most basic hydrologic landscape elements
- Topography dictates the flow of water across the landscape
- Flowing water sculpts the landscape



- This synergy is at the heart of much hydrologic modeling relating to questions of runoff generation important for flooding and water resources
- Representing hydrologic processes at high resolution is important to help solve these problems

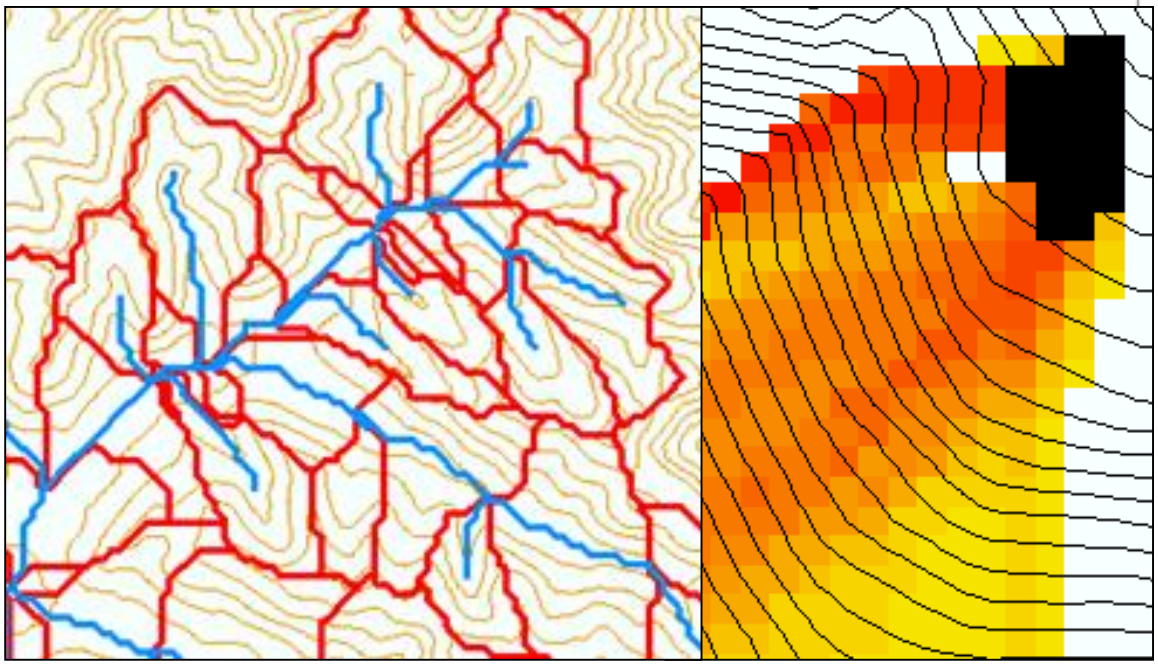
TauDEM is software for deriving hydrologically useful information from Digital Elevation Models



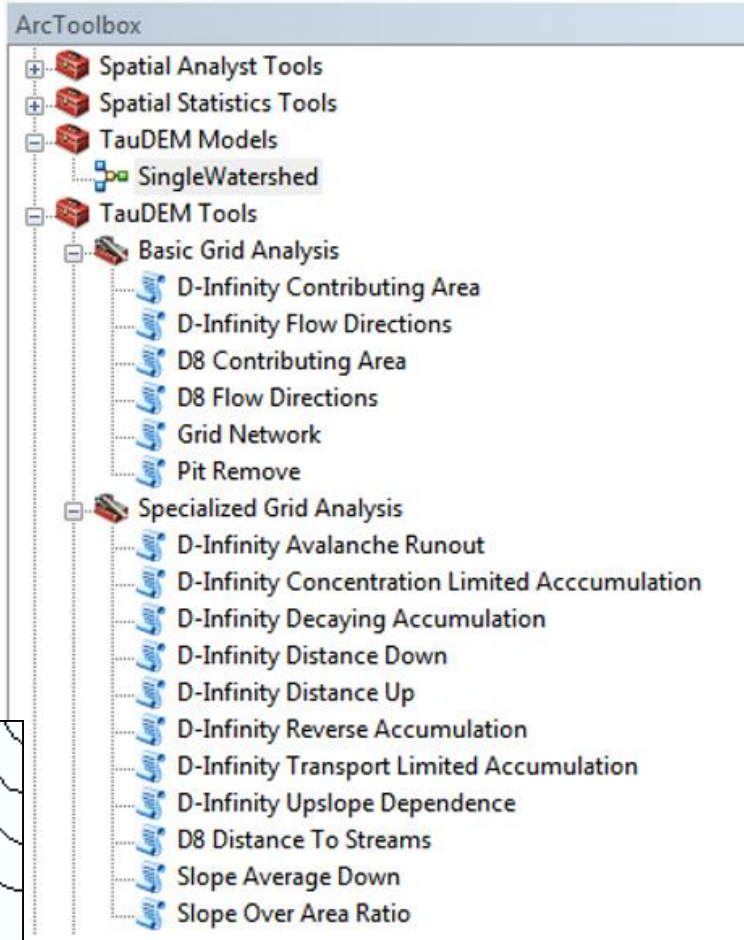
http://hydrology.usu.edu/taudem/

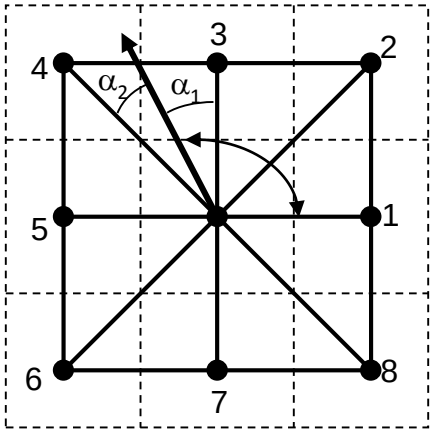
TauDEM

- Stream and watershed delineation
- Multiple flow direction flow field
- Calculation of flow based derivative surfaces
- MPI Parallel Implementation for speed up and large problems
- Open source platform independent C++ command line executables for each function
- Deployed as an ArcGIS Toolbox with python scripts that drive command line executables



http://hydrology.usu.edu/taudem/





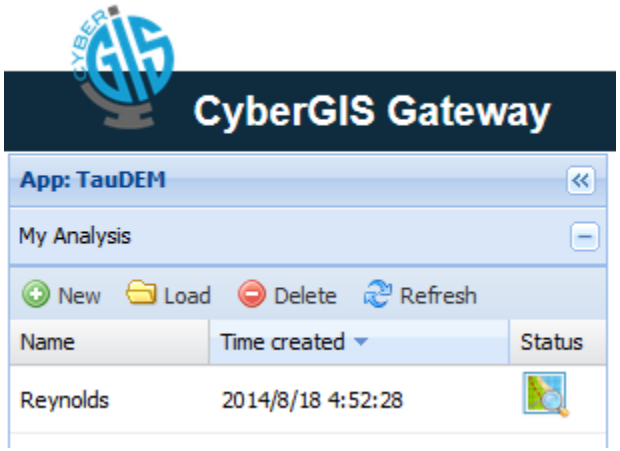
Using TauDEM today requires

- Expertise in Hydrologic DEM analysis
- The software
  - ArcGIS licenses (for ArcGIS plugin)
  - The ability to install software
  - TauDEM command functions with MPI installation
  - Compilation for other platforms  
- Sufficient Hardware (RAM and Disk)
- The data (GDAL formatted rasters with consistent grid size and spatial reference)

Moving TauDEM to the cloud

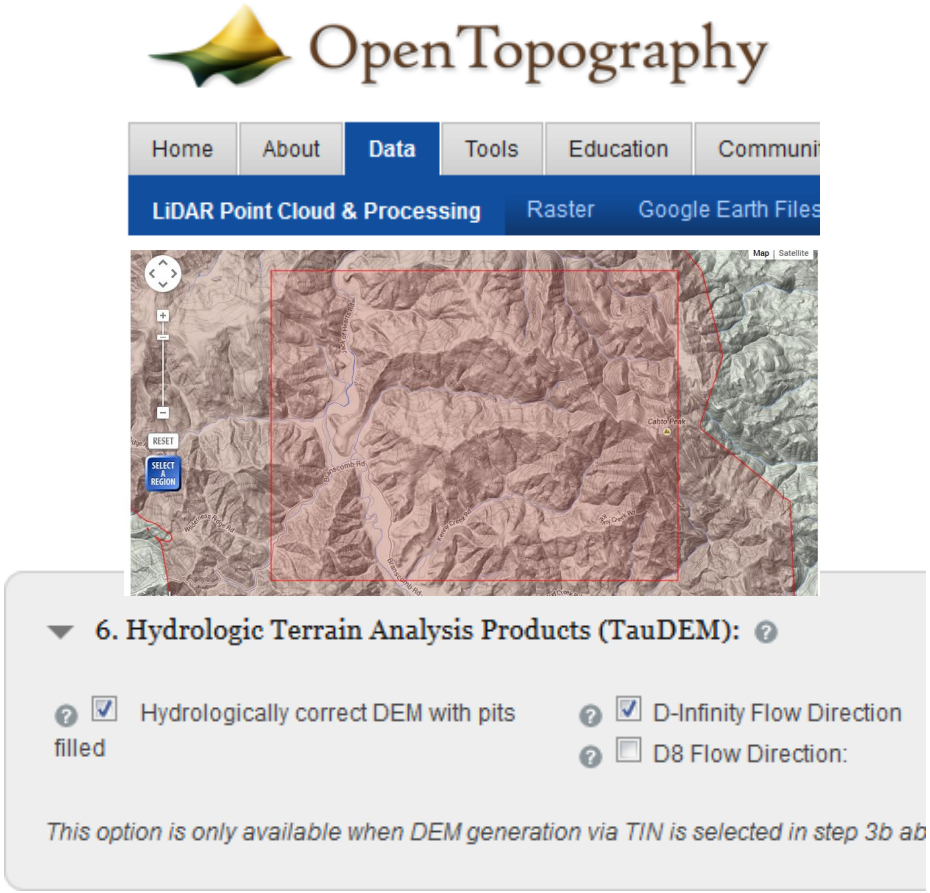
CyberGIS

http://gateway.cigi.illinois.edu/



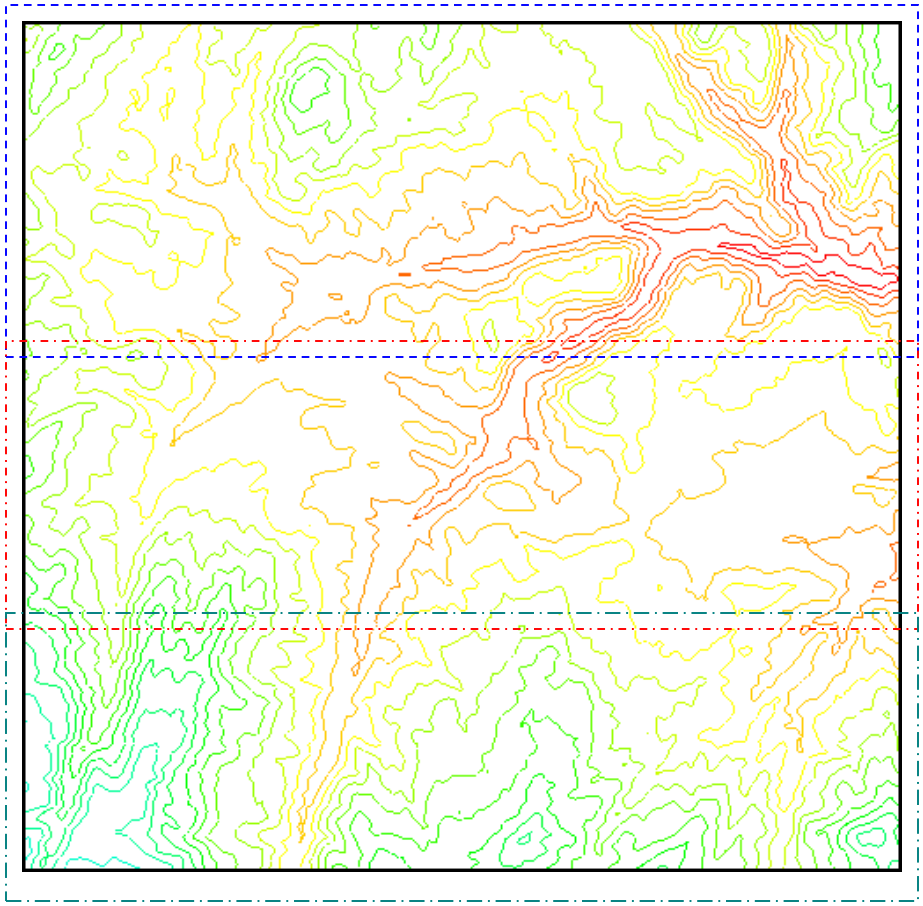
Open Topography

www.opentopography.org



TauDEM Parallel Approach

- MPI, distributed memory paradigm
- Row oriented slices
- Each process includes one buffer row on either side
- Each process does not change buffer row
- Improved runtime efficiency
- Capability to run larger problems



Open Topography Data and Product Selection

Select Data Product: Point Cloud Download & Processing | Raster | Google Earth Imagery File | Point Cloud Bulk Download

1a. Select area of data to download or process:

Horizontal Coordinates: UTM z10 N NAD83 (CORS96) [EPSG: 26910] - **Vertical Coordinates**: Ellipsoid (GRS80)

Data Selection Coordinates: ☐ Manually enter selection coordinates (in the horizontal coordinate system listed above)

X_{min} = 442392.942 Y_{min} = 4390197.763 X_{max} = 450575.817 Y_{max} = 4396426.705

The selection area contains approximately 8,504,000 points.

1b. Choose Return Classification: ☑ Ground ☐ Unclassified

LABS

6. Hydrologic Terrain Analysis Products (TauDEM):

☑ Hydrologically correct DEM with pits filled

☑ D-Infinity Flow Direction

☐ D8 Flow Direction:

☑ D-Infinity Specific Catchment Area

☐ D8 Contributing Area

☑ Topographic Wetness Index

This option is only available when DEM generation via TIN is selected in step 3b above.

Open Topography Result

Job Id	Dataset	Title	Submission	Completion	Duration	Num points	Final Status
1408355339888190271563 3	CA04_Power	Eel River TauDEM analysis	2014-08-18 02:49:00	2014-08-18 02:58:39	579 secs	7,647,196	Done

Download Job Metadata:

metadata-1408355339888190271563 3.txt

Download Job Results

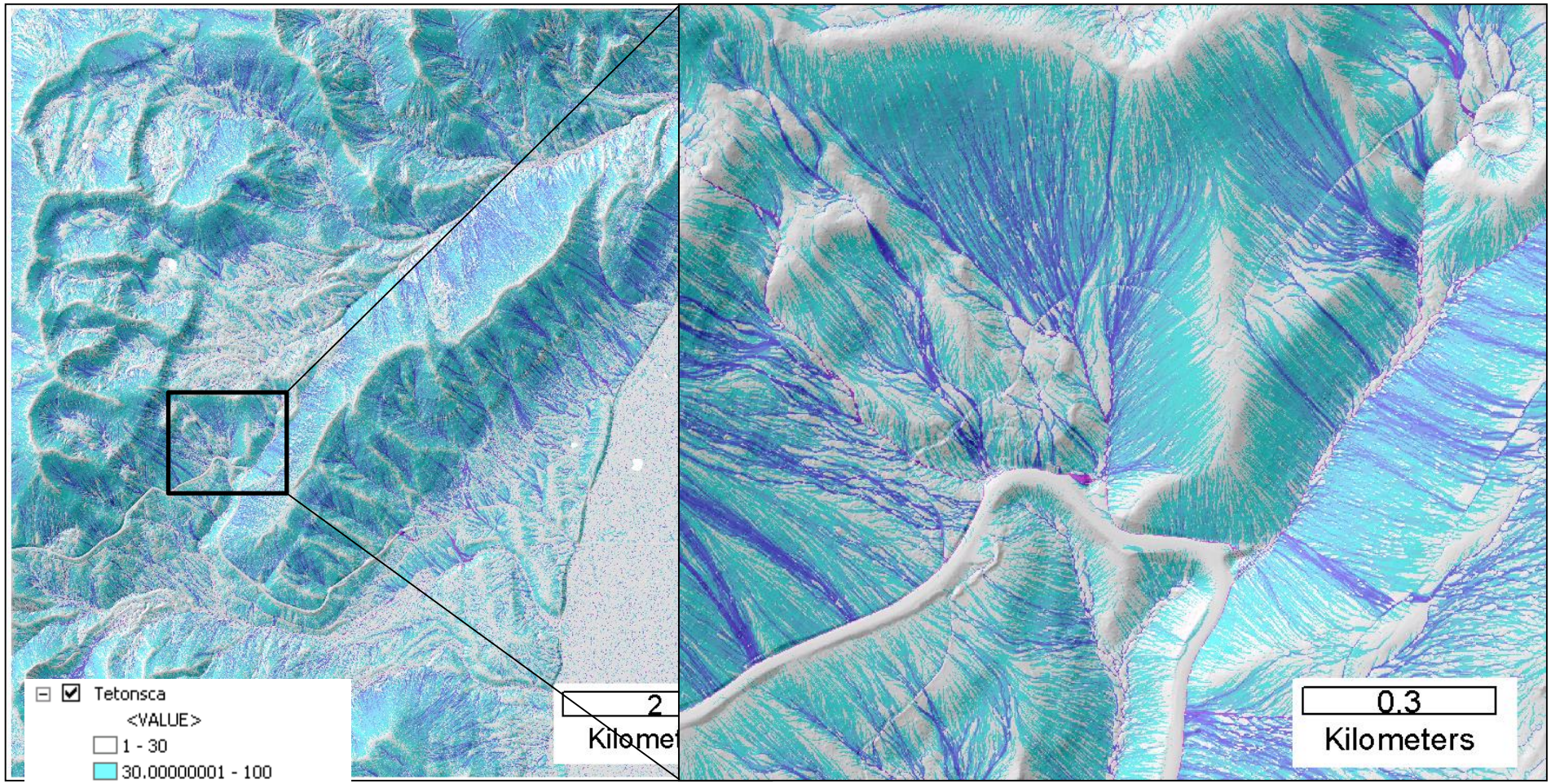
Point Cloud Results:
Download point cloud data in LAS format: points.las

DEM Results:
Download DEM (Local Gridding): dems.tar.gz
Download DEM (TIN): dems.tar.gz

Derivative Product Results:
Download Hillshade & Slope Products (Local Gridding): viz.tar.gz
Download Hillshade & Slope Products (TIN): viz.tar.gz

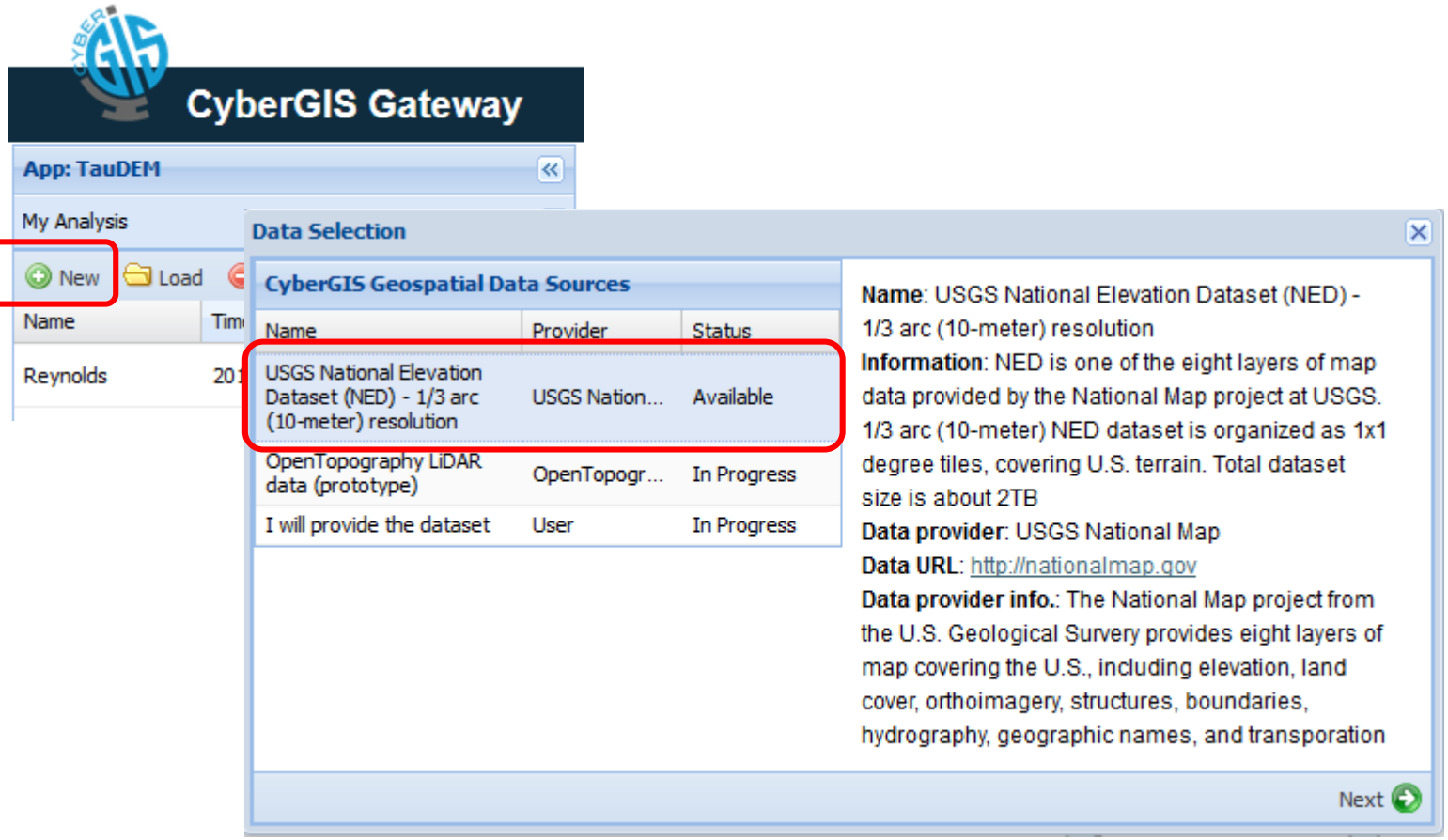
TauDEM Product Results:
Download PitRemove file: pitRemove.tar.gz
Download D-Infinity: Flow Direction file: dinfFlowDirection.tar.gz
Download D-Infinity: Slope file: dinfSlope.tar.gz
Download D-Infinity - Area Contribution file: Dinfarea.tar.gz
Download Topographic Wetness Index file: TWI.tar.gz

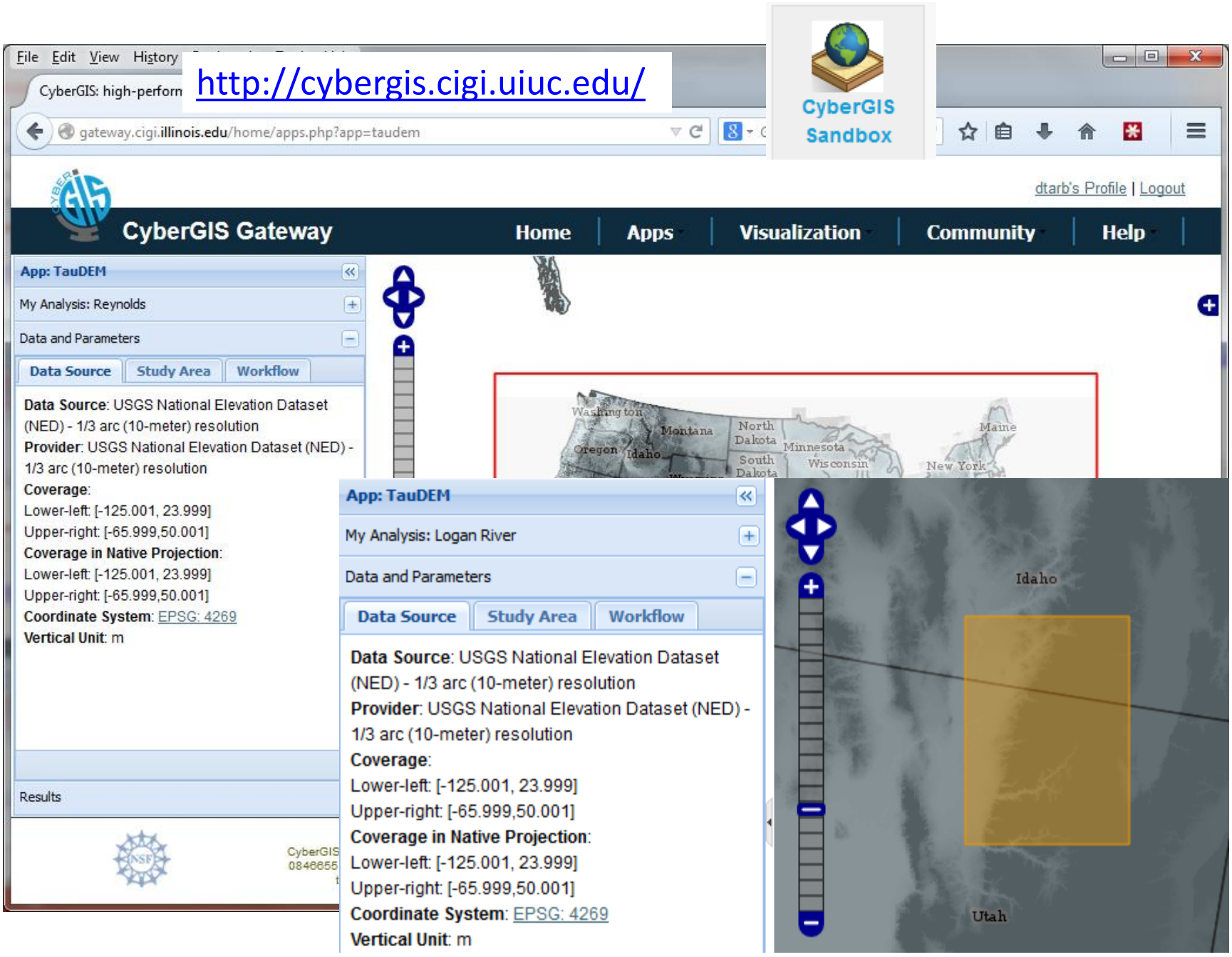
Contributing area from D-Infinity



CyberGIS TauDEM App

http://gateway.cigi.illinois.edu/




http://cybergis.cigi.uiuc.edu/
CyberGIS Sandbox
CyberGIS Gateway
Home
Apps
Visualization
Community
Help
dtarb's Profile | Logout
App: TauDEM
My Analysis: Reynolds
Data and Parameters
Data Source
Study Area
Workflow
Data Source: USGS National Elevation Dataset (NED) - 1/3 arc (10-meter) resolution
Provider: USGS National Elevation Dataset (NED) - 1/3 arc (10-meter) resolution
Coverage:
Lower-left: [-125.001, 23.999]
Upper-right: [-65.999,50.001]
Coverage in Native Projection:
Lower-left: [-125.001, 23.999]
Upper-right: [-65.999,50.001]
Coordinate System: EPSG: 4269
Vertical Unit: m
Results
App: TauDEM
My Analysis: Logan River
Data and Parameters
Data Source
Study Area
Workflow
Data Source: USGS National Elevation Dataset (NED) - 1/3 arc (10-meter) resolution
Provider: USGS National Elevation Dataset (NED) - 1/3 arc (10-meter) resolution
Coverage:
Lower-left: [-125.001, 23.999]
Upper-right: [-65.999,50.001]
Coverage in Native Projection:
Lower-left: [-125.001, 23.999]
Upper-right: [-65.999,50.001]
Coordinate System: EPSG: 4269
Vertical Unit: m
Idaho
Utah

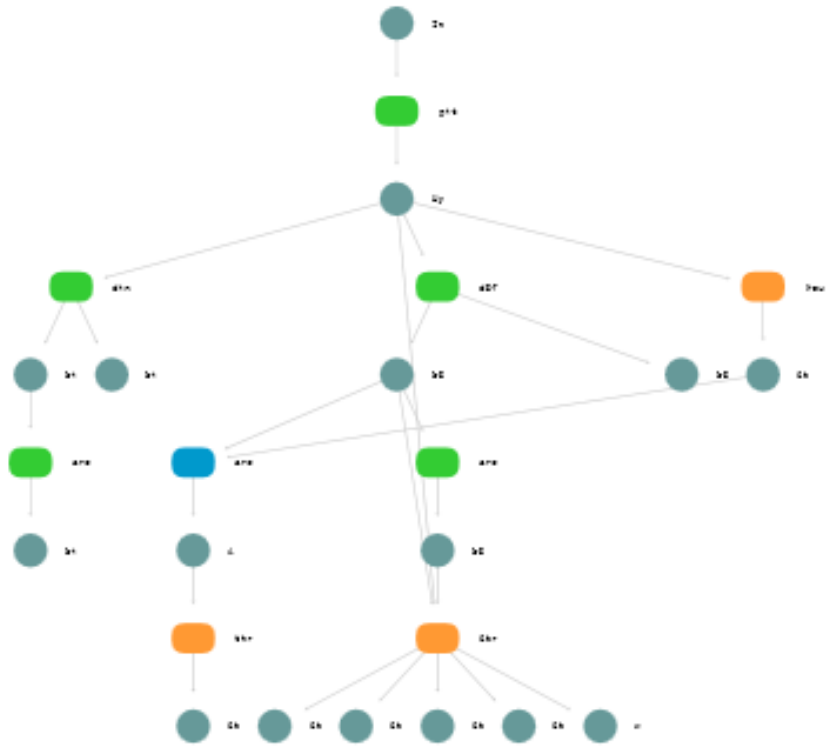
Select the products you want

TauDEM Workflow Wizard

Select Products

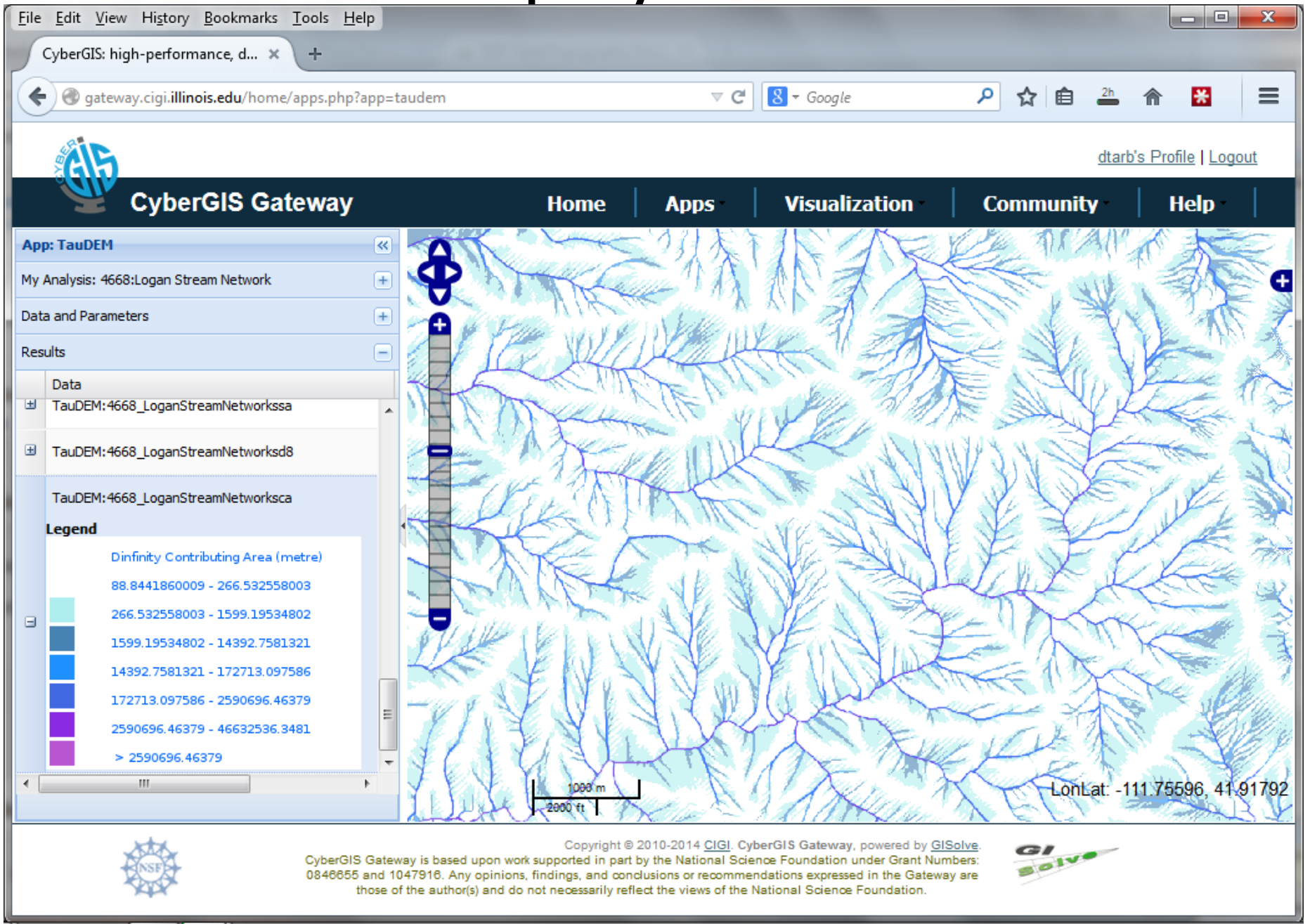
Filter

	Name	RID
	Common TauDEM Products	
☐	Hydrologically Conditioned Elevation Grid	1
☐	D8 Flow Direction	3
☐	D8 Slope	2
☐	D8 Contributing Area	6
☐	Dinfinity Flow Direction	5
☐	Dinfinity Slope	4
☑	Dinfinity Specific Catchment Area	7
☐	Contributing Area Stream Raster	14
☑	Peuker Douglas Stream Raster	15
☑	Stream Network And Subwatersheds	21
☐	Gage Subwatersheds	27
☐	Topographic Wetness Index	22
	Specialized TauDEM Products	
☐	Grid Strahler Order	8
☐	Grid Path Length	9
☐	Grid Total Length	10
☐	D8 Flow Accumulation Options	11



The wizard configures the sequence of functions to run to get the result

Results displayed in browser



Advancing Data Services for Modeling and Analysis

Assumptions

1. GIS and hydrologic modelers have to learn and become comfortable using a modern scientific programming language (e.g. Python or R)
2. Modeling is data intensive (large datasets from a range of sources) demanding more data and computing resources than is in most PC's
3. Reproducibly installing and configuring models on different platforms is a challenge
4. Should not have to become expert in HPC systems and learning them is a barrier to using HPC and research with big data and computationally intensive models

Computation via Python Client calling Data and Modeling Services

Input

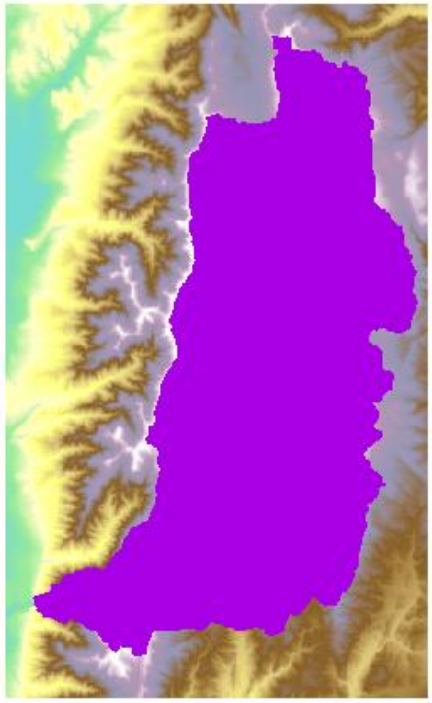
modeldemo.py × demo.py ×

```
from hydrogate import Client
client = Client()
client.subset_dem(left_top_x=410000, left_top_y=4682453
# Previously created shapefile
outlet_shapefile_url = "http://129.123.41.158:8080/dem/
subset_dem_request = client.get_most_recent_request(ser
print subset_dem_request.file_path
input_raster_url = subset_dem_request.file_path
client.generate_watershed_raster(input_raster_url_path=input_raster_url,
                                 outlet_shapefile_url_path=outlet_shapefile_url,
                                 save_as=r'C:\Users\dtarb\Desktop\HydroGateDemo\WS_Logan.tif')
```

Python session on desktop but data and analysis on server

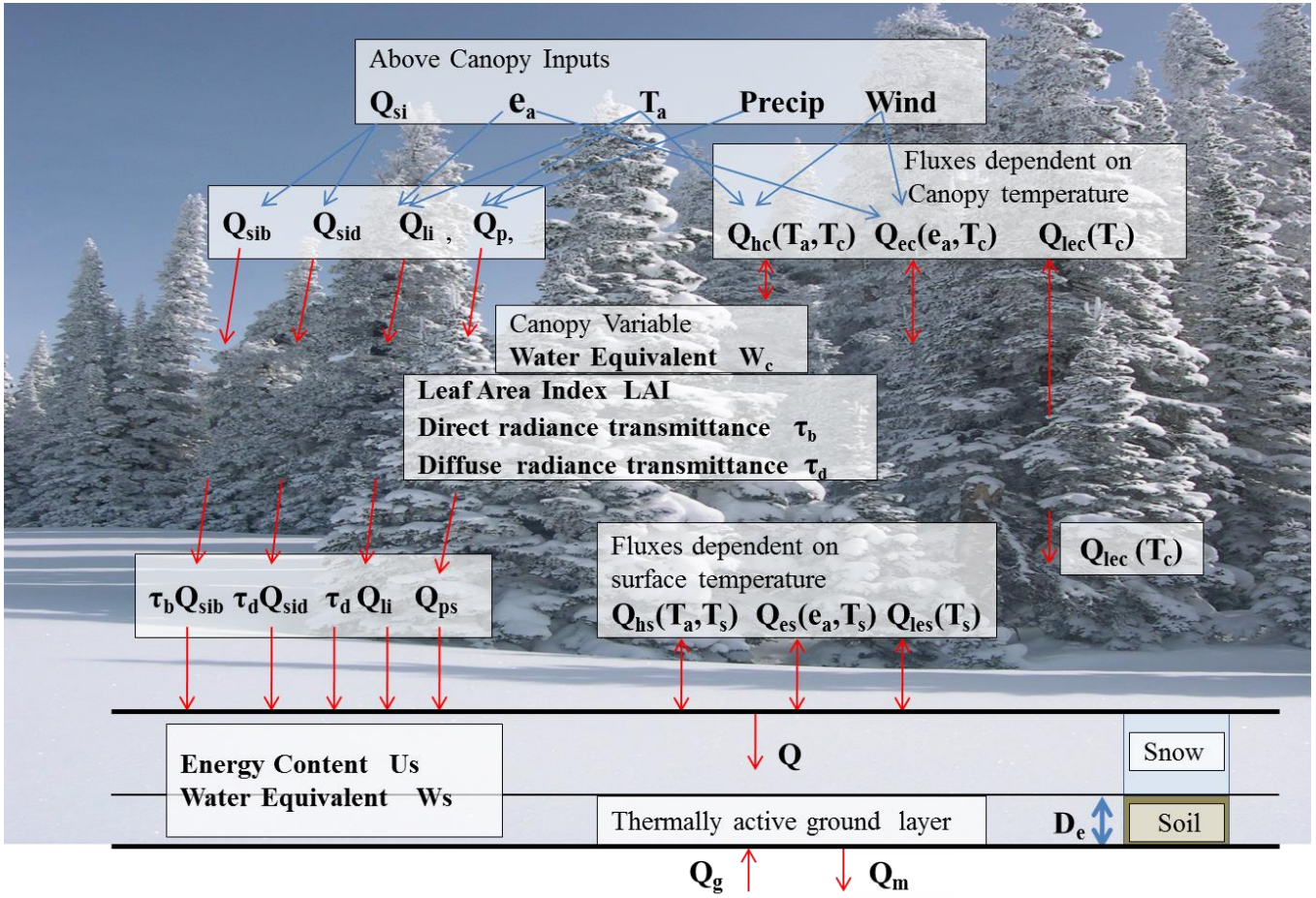
Result

```
C:\Python27\ArcGIS10.2\python.exe C:/Users/dtarb/Desktop/HydroGateDemo/modeldemo.py
subset_dem execution was successful.

Output file URL path:http://129.123.41.158:8080/dem/dem3937079318519734987.tif.zip

{
    "Output file path": "http://129.123.41.158:8080/dem/dem3937079318519734987.tif.zip",
    "Service status": "success",
    "Request time": "2014-11-16 00:26:20.548000",
    "Service name": "subset_dem",
    "Service ID name": "",
    "Service ID value": ""
}
http://129.123.41.158:8080/dem/dem3937079318519734987.tif.zip
generate_watershed_raster execution was successful.

Output file URL path:http://129.123.41.158:8080/dem/dembfa12a4fdf924fa6b0896f892eb837c8WS.tif.zip

Downloaded file saved successfully at:C:\Users\dtarb\Desktop\HydroGateDemo\WS_Logan.tif
```



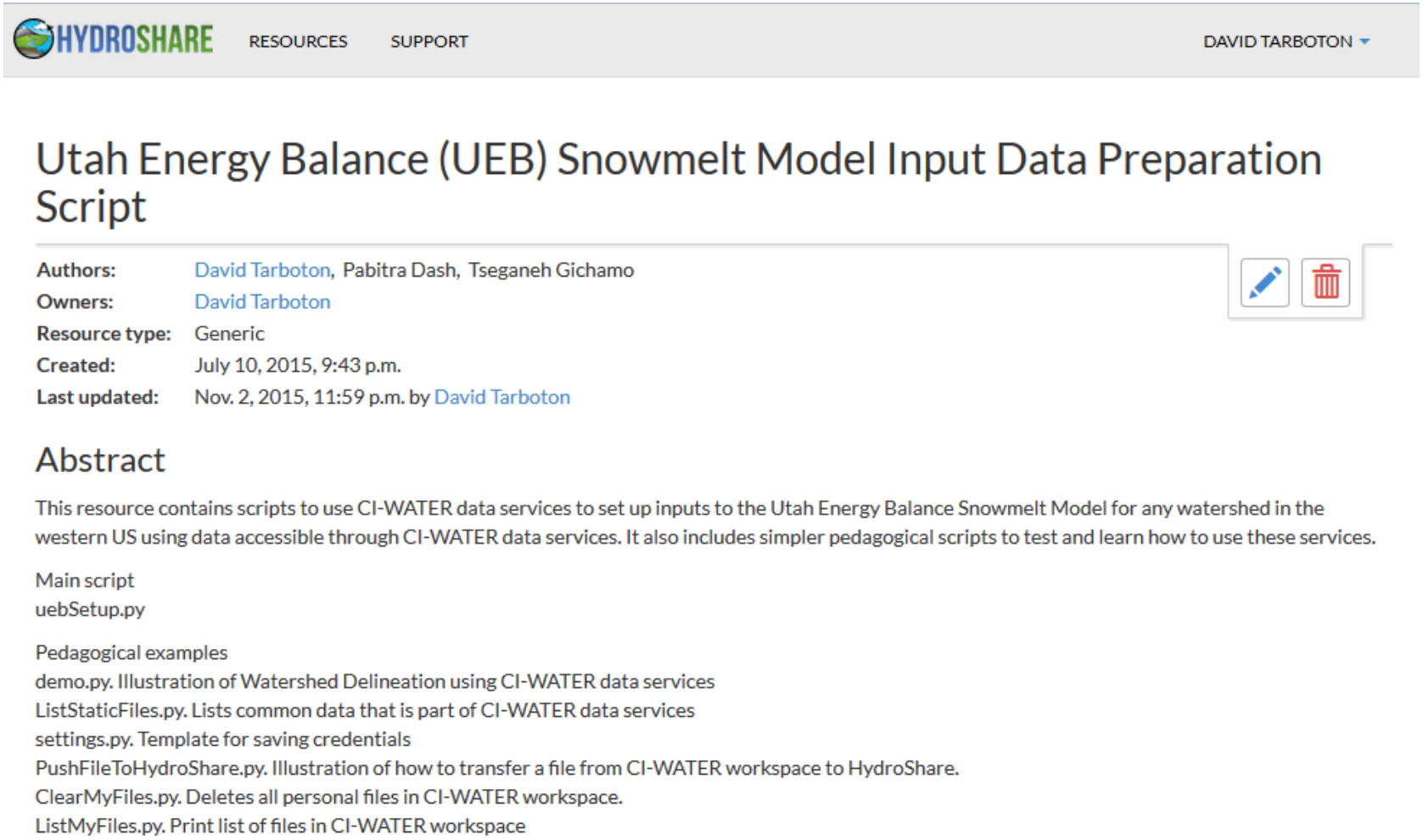
Utah Energy Balance Snowmelt Model

Used to address what are the impacts of land cover change on watershed snowmelt inputs

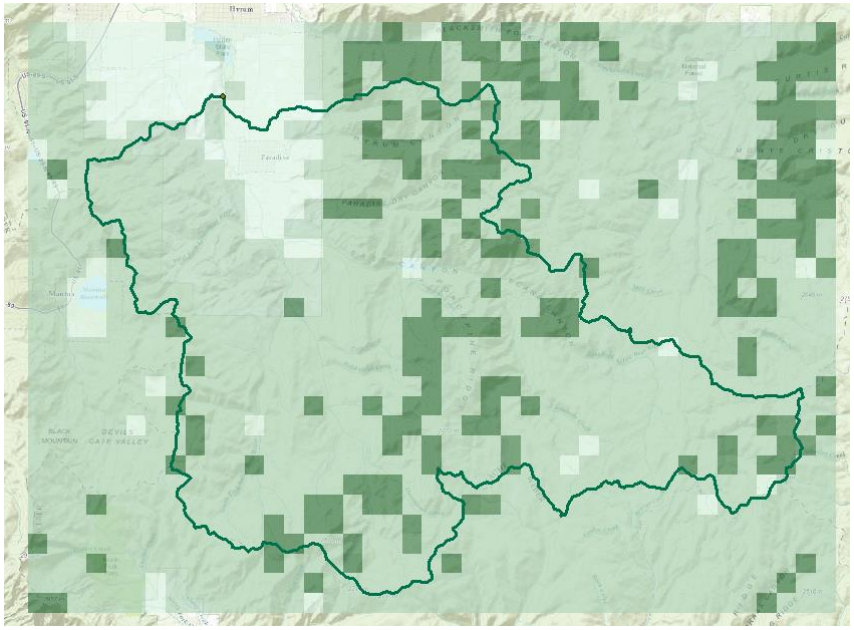


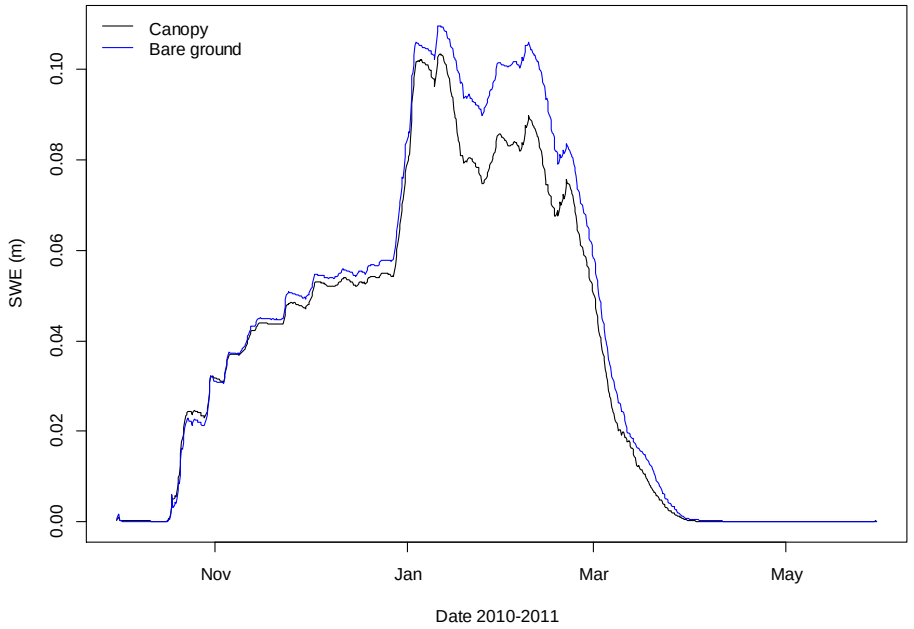
Mahat, V. and D. G. Tarboton, (2012), "Canopy radiation transmission for an energy balance snowmelt model," Water Resour. Res., 48: W01534, http://dx.doi.org/10.1029/2011WR010438.

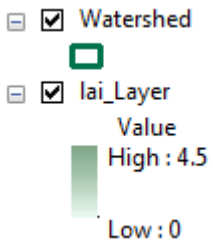
Example preparation of inputs for UEB using HydroDS Services



Use UEB to examine Sensitivity of SWE to Canopy removal





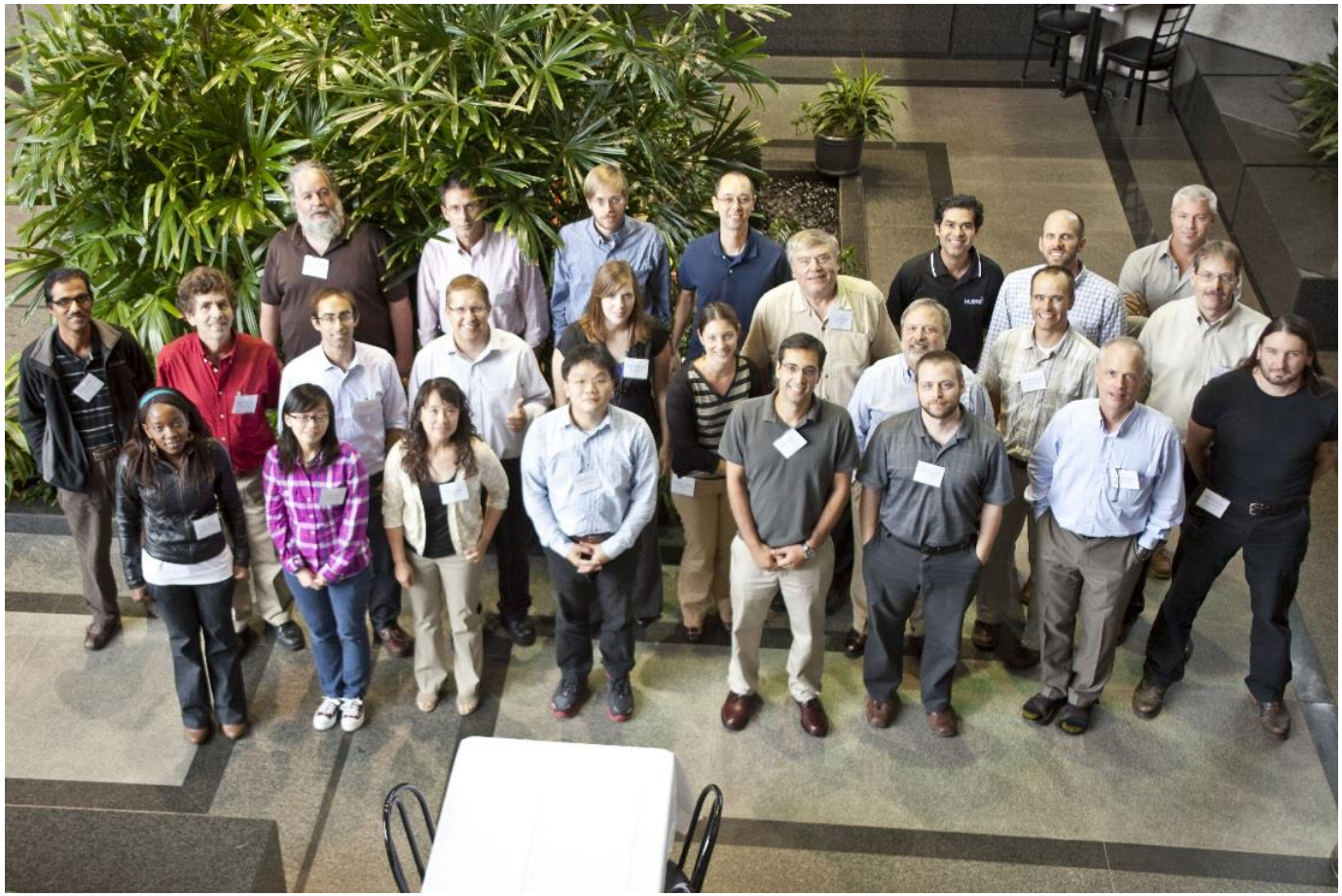


Summary

1. A new, web-based system for advancing model and data sharing
2. Access multiple types of hydrologic data using standards compliant data formats and interfaces
3. Flexible discovery functionality
4. Model sharing and execution
5. Facilitate and ease access to use of high performance computing
6. Social media and collaboration functionality
7. Links to other data and modeling systems

Thanks to the HydroShare team!

- USU
- RENCI/UNC
- CUAHSI
- BYU
- Tufts
- UVA
- Texas
- Purdue
- SDSC



HYDROSHARE

http://www.hydroshare.org

NSF
OCI-1148453
OCI-1148090
2012-2017